ResumeRDF: Expressing skill information on the Semantic Web

Uldis Bojārs, John G. Breslin

Digital Enterprise Research Institute,
National University of Ireland, Galway
[uldis.bojars; john.breslin]@deri.org

Abstract

This paper explores the ResumeRDF ontology [1][2] and its potential application to the tasks of ExpertFinder initiative, including describing people's resumes and skill sets. We consider DOAC [3] as another ontology which can be used skill information, compare their characteristics and describe directions for future development. Second part of this paper looks at discovery of skill information on the Semantic Web which becomes important as soon as skill information starts to get published.

1. Introduction

The Semantic Web is becoming more common. Resource Description Framework (RDF), one of foundations of the Semantic Web, facilitates integration of heterogeneous information from distributed information systems, such as different enterprise information systems or data sources spread throughout the Web. Availability and creation of vocabularies to describe business processes is an important success factor for the Semantic Web, and human resources is one of the areas where such vocabularies are needed.

A resume or Curriculum Vitae (CV) is a written record of person's education and employment, and is usually submitted when searching or applying for a job. Resume usually include description of person's work experience and skills.

People's resumes and skill profiles are a kind of information that is of a particular relevance to the ExpertFinder initiative which aims to develop vocabularies, their rule extensions, best practices and recommendations towards standardization of metadata that would enable computer agents to find experts on particular topics.

By collecting information from profiles people have made available on their home pages and online community sites one could a get an insight on people who are experts in their area. Provided that skills are described in sufficient detail, include quantitative measures and use a common classification scheme we should also be able to compare and rank the levels of expertise.

The rest of this paper is organized as follows: Section 2 looks at properties of two ontologies that can be used to describe skills resume information. Section 3 concentrates on discovering skill information on the Web. Section 4 contains a summary which concludes the paper.

2. Ontologies for Skill and Resume Information

ResumeRDF is an ontology developed in order to express on the Semantic Web information contained in a resume, such as business and academic experience, skills, publications, certifications, etc. This information is expressed using 2 namespaces:

- http://purl.org/captsolo/resume-rdf/0.2/cv - Resume ontology
- http://purl.org/captsolo/resume-rdf/0.2/base - Property value taxonomy

Resume described by this ontology consists of information about the person, its work and academic experience (education), skills, courses and certification, publications, references and other information.

Description of a Career (DOAC) is a vocabulary to describe professional capabilities of workers. This ontology is defined using the namespace:

- http://ramonantonio.net/doac/0.1/

We will be exploring these vocabularies throughout this paper. ResumeRDF vocabulary was designed by one of the authors, therefore more information about it is available to us and we are able to use this study to make decisions about its future development.

2.1. General comparison

Both vocabularies are able to describe information about a person, her experience and skills, and have parts that are

common and also some important differences. Overall comparison:

- DOAC contains 15 classes and 17 properties and is geared towards describing person's skills and experience (created in 2005)
- ResumeRDF contains 16 classes and 73 properties and is aimed at describing full contents of a Curriculum Vitae (created in 2002)

ResumeRDF is designed with querying in mind and is able to capture a larger amount of semantic information, e.g. Organisation is a class on its own, is related to person's work history using a cv:employedIn property, can be given URIs and assigned additional properties. Current job can be described using cv:isCurrent and therefore easily identified. A more detailed description of these properties is not within the scope of this paper and is available in [1].

DOAC describes most of the experience information, e.g., organisation names as text literals. This adds less overhead when creating data instances but can be prone to typing mistakes and limit ability to find persons who have worked in the same company or studied in the same university. An area that is better developed in DOAC is separate subclasses for education, e.g. PrimarySchool and SecondarySchool. According to the specification DOAC is compatible with the European Curriculum.

Both use start date and end date to describe the time-frame of work and academic experience.

2.2. Personal information

Both ontologies use FOAF (Friend-Of-A-Friend) [4] to express information about persons and their relationships and thus may extend personal FOAF profiles. Resume RDF also recognizes that FOAF alone is not sufficient for all the personal details described on a resume and suggests to use vCard [5] for the missing contact information.

Some personal information is not covered by either FOAF or vCard. For this information ResumeRDF introduces new properties, e.g. cv:birthPlace, cv:hasCitizenship or cv:maritalStatus.

When describing personal information privacy considerations, applicable laws and regulations have to be taken into account and some of these properties (e.g., marital status) may not be allowed to be requested in certain countries. At the same time there may be countries where putting this information in a resume is a common practice. Therefore these properties are included in the ontology and it is the system design time when choices need to be what subset of classes and properties to use.

2.3. Skill information

Skills in ResumeRDF are described with cv:skillName, cv:skillLevel (0-5), cv:skillLastUsed and cv:skillYearsExperience. This allows to quantify of skill levels. While initial design of the ontology uses literals to describe skills, [1] suggests to use URIs instead to avoid ambiguity of skill identification and enabling automatical skill matching. It has a subclass for language skills and allows to specify if a person has a valid driver's license.

DOAC specifies 4 different kinds of skills: language, social, organisational and computer. Specification shows examples of usage of doac:LanguageSkill which is well defined using ISO-639-1 codes for languages and skill levels for proficiency. Skill levels shown in examples are not defined in the ontology but we assume that this will be corrected in future versions of the ontology. There are no further details in the ontology or specification on how other types of skills are defined.

We can assume that both ontologies allow to describe skills and their levels but more details on describing other types of skills in DOAC to be able to compare them.

2.4. Taxonomies for resume information

An important question for expert finding is to agree on taxonomies or property values to use for skill identification. Using a distributed list of URLs of web sites describing theses skills or programming languages is not practical because in order to provide a list of all possible property values developers would need to search the whole Web and identify all possible URLs which might be used to describe skills.

This can be addressed by choosing a set of taxonomies or along with URIs of skills in them to define skills. In order to help creation of resume skill profiles helper services may be created which keep track of most used skill identifiers and suggest values to fill in. A potential extension of such services can be a social software site similar to del.icio.us where people can register new skills in the case if existing choices do not contain the skills they want to describe.

There are other kinds of resume information that have a set of values to choose from. Some of these sets are more constrained in the value they can take and can therefore be defined within an ontology. Taxonomy module of ResumeRDF (base.rdfs) defines taxonomies that can be used to describe resume information, e.g., a career level (Student, EntryLvl, ..., SeniorExecutive) or a job type (Employee, Contractor, Intern).

3. Discovery of Skill Information on the Web

People may start to use these and other ontologies to describe their skills on the web but there is another important task for expert finding - discovery and aggregation of skill information. A simple and efficient solution is to adapt existing social networking and online community sites to allow people to describe their skills using one of these formats. These sites are able to make massive amounts of data available but the question to be answered is how to motivate them to do this and what benefits does it create for them.

Developer community sites such as Advogato are well suited for this purpose because they already allow to describe persons skills and trust metrics.

Another potential source of software skill information is software projects which people have created. If we know that a person has created a highly ranked software project in, e.g., Python programming language, we can assume that he has a considerable level of skills in Python and associated technologies (e.g., version control or web services) that are required to develop this software project. This information could be sourced from open source software development sites such as SourceForge.net if it was made available in a machine readable form, e.g., using DOAP [6] and SIOC [7] vocabularies.

In a distributed publication scenario where people publish their skill information on personal web pages we need a different approach. Information is distributed throughout the Web and needs to be discovered and aggregated in order to be useful. O'Reilly defines the network effect of user contributions as one of the key aspects for success in today's dynamic web or Web 2.0 [8] and it can be equally useful in resource discovery on the Semantic Web. PingTheSemanticWeb is a web service that collects notifications about presence or updates to RDF data on the web and provides these URLs to other Semantic Web applications which are able to use this data. Combine it with a Semantic Radar extension for Firefox which lets users to discover RDF data on web pages which they are browsing and can notify PingTheSemanticWeb about presence of RDF data and the result is a platform for collaborative discovery of data, including resumes and skill profiles, which users may contribute to while browsing the web as usual.

4. Summary

This paper looked at two ontologies developed to describe skills and resume information - DOAC and ResumeRDF. Both ontologies have their strengths and weaknesses. ResumeRDF is able to describe additional semantic relations and provide a richer source of information but it adds another level of indirection between a person and its skills which can be considered as its downside. DOAC has a simpler structure and mainly uses literals to express information about work experience but this can also be considered as a benefit because it makes it simpler to authoring resume data.

We can conclude that both ontologies should be developed further and can learn from each other. Analysis performed in this paper gives us a direction for future development of the ResumeRDF ontology which we consider to be a future work. As a first step of this work we have published an updated ResumeRDF specification at [2].

Discovery of information is as important as creating it. We looked at two approaches to collect this information using large community sites and using user-driven collaborative discovery of RDF information. Both approaches can be combined and provide as good source of skill information.

With skill information described on the Web and aggregated from large community sites or individual web pages the next task is to create applications that make use of this information and find best experts in a given area. We expect the ExpertFinder initiative to give significant input to this area and be one of real use case s for the Semantic Web.

References

1. Uldis Bojārs. *Extending FOAF with Resume Information.* 1st Workshop on FOAF, Social Networks and the Semantic Web, 2004.
2. Uldis Bojārs. *ResumeRDF Vocabulary Specification*. Available at: http://rdfs.org/resume-rdf/
3. Ramon Antonio Parada. *DOAC Vocabulary Specification.* 2006-07-08. Available at: http://ramonantonio.net/doac/0.1/
4. Dan Brickley, Libby Miller. *FOAF Vocabulary Specification.* Available at: http://xmlns.com/foaf/0.1/
5. Renato Ianella. *Representing vCard Objects in RDF/XML.* W3C Note 22 February 2001. Available at: http://www.w3.org/TR/vcard-rdf
6. Edd Dumbill. *XML Watch: Describe open source projects with XML.* IBM DeveloperWorks, 2004-02- 26. http://www-128.ibm.com/developerworks/xml/library/x-osproj.html
7. J. G. Breslin, A. Harth, U. Bojārs, and S. Decker. *Towards Semantically-Interlinked Online Communities.* In 2nd European Semantic Web Conference, pages 500–514, May 2005.
8. Tim O'Reilly. *What Is Web 2.0: Design Patterns and Business Models for the Next Generation of Software.* 2005-09-30. http://tim.oreilly.com/news/2005/09/30/what-is-web-20.html